Application of a computer controlled XY tablet for paper and pencil tests in neuropsychological assessment

Computers in Psychology

Editors: F.J. Maarse
L.J.M. Mulder
W.P.B. Sjouw
A.E. Akkerman

Methods, Instrumentation and Psychodiagnostics

Application of a computer controlled XY tablet for paper and pencil tests in neuropsychological assessment

P. Houx[1], N. Brand[2], and J. Jolles[1]

Abstract

The present study describes the application of an XY tablet in combination with a paper and pencil variant of the Memory Comparison Task (Sternberg, 1969). This information processing test is frequently used in the differentiation of particular cognitive processes. The data indicate that implementation on an XY tablet yields essentially the same results as when the normal test version is used. In addition, many more aspects of the subject's information processing can be reliably assessed in the computer-controlled version. The advantages of paper and pencil tests and computerized testing seem to be combined by the procedure proposed in the present study.

1. Introduction

The use of paper and pencil tests is widespread in neuropsychological research and assessment (see e.g. Lezak, 1983). The strength of these tests is their short duration, simple instruction and the relative ease of obtaining the test material. In addition, many different tests are available enabling the measurement of several aspects of cognitive performance. The fact that the test sheet and the marks on it yield a hard copy of the performance can also be considered an advantage. In the traditional paper and pencil test, a test sheet which contains a number of stimuli is presented to the subject. Characteristically, the subject is instructed to respond to particular stimuli by making a mark with the pencil on the paper. Examples of such tests are the Trail Making Test (see Lezak, 1983, for a discussion) in which digits and/or letters are to be connected, Embedded Figures Tests (Thurstone, 1944) in which particular stimuli have to be matched and indicated, and the Utrecht Memory Comparison Task (Jolles, Hijman & Gaillard, 1982; Brand & Jolles, 1987) which is based upon the paradigm described by Sternberg (1969,1975). This task requires the subject to mark those stimuli which can be matched to stimuli in short term memory.

Paper and pencil tests have their disadvantages as well. In these tests, performance is measured by comparing the time needed per test sheet to norms. Unfortunately, this is a crude way of measuring since all stimuli are presented together. As a result, the total time and the total number of errors per test or test page are the only parameters that yield quantitative data. The variability of performance cannot be measured since the duration of a single response (i.e. the time lapse between two pencil strokes, or the inter-response time) cannot be determined. An additional disadvantage refers to the fact that the subject gets feedback on his performance through the pencil strokes that have been made. This is sometimes undesirable as the subject may learn from his/her own performance. In addition, a test may become easier when the subject can use information which is indicated on the paper by the pen marks.

The advent of computer technology may enable the development of new methods of psychological assessment that combine the advantages of the paper and pencil tests to the more sensitive detection and measurement of computer aided assessment. The present study explores the possibilities of a computer controlled XY tablet (graphic digitizer) for this purpose. The only difference that can be noted by the subject is the nature of the pencil, because the pen used is connected to the computer by a wire. Theoretically, this method of testing combines the necessary simplicity and comprehensibility to the sensitivity and accuracy of computer controlled research methods. Furthermore, it enables the measurement of individual marks of the pen on the tablet and the determination of a measure of individual responses without interfering with the performance on the orthodox paper and pencil test.

This paper describes the use of a recently developed paper and pencil test

[1] State University of Maastricht, Clinical Psychiatry, Department of Neuropsychology and Psychobiology, P.O. Box 616, 6200 MD Maastricht.

[2] State University of Utrecht, Department of Clinical Psychology, Heidelberglaan 1, 3584 CS Utrecht.

on an XY tablet in combination with a microcomputer. The questions addressed in this study were: 1) whether there is a correspondence between the results obtained with the original test and those obtained in the computerized version, and 2) whether the measurement of individual stimuli adds relevant information.

2. Materials and Methods

2.1. Test description

The paper and pencil task under investigation in the present study is the Utrecht Memory Comparison Task (UMCT; Jolles, Hijman and Gaillard, 1982). This is a variant of the memory comparison paradigm developed by Sternberg (1969, 1975). According to Sternberg, several stages of information processing can be assessed independently by analysing the relation between task factors in their effect on reaction time (RT). It has thus been shown that a linear relation exists between the time that is needed to compare a series of items with similar items in memory, and the number of items held in memory (the memory load). In a graphical representation of this relation, the intercept is taken to be a measure of the rate of perception and motor response. The slope of this RT-setsize function is a measure of the memory scanning process. Sternberg proposed the following underlying processing stages:

A. an initial encoding stage,
B. a memory comparison stage (scanning and retrieval),
C. a binary decision stage, and
D. a stage of motor organization and response execution.

The UMCT is a paper and pencil version of the orthodox procedure with computer screens or slide projectors. It has several advantages such as ease of presentation even in clinical settings, and several studies have shown that the task meets all the theoretical requirements that have been found for the orthodox task. That is, there is a linear relation between setsize and reaction time. Furthermore, the task has proved to be of importance in use in clinical settings (Jolles, 1985; Brand and Jolles, 1987). Additional advantages refer to the fact that the task lends itself especially to manipulation of the task parameters, by using a set of different test sheets.

The task has three parts. In the practice part the symbol "%" is administered as memory set, on a separate sheet. The subject is asked to search for this symbol on the test form, consisting of 144 characters typed in a 12 x 12 matrix, 4 spaces apart from each other. One sixth of these characters consists of items from the memory set, the rest are capital consonants. In part two, three different forms are presented with 1, 2, and 4 digits as memory set. Part three consists of four sheets (memory set sizes 1 through 4 letters). 'Distractors' are always letters (see also Figure 1).

2.2. Hardware

For the computerized administration of paper and pencil test, the following hardware was used (Figure 2): The Apple Graphics Tablet (1) and the pen (2) are connected by wires with a Basis-108 (3), which is compatible to the Apple IIe, with a normal monochrome monitor. A matrix printer (4) serves to show the results of the analysis of the provisional data. This set-up has been developed by Brand and Jolles (1987).

2.3. Software

The computer program for the test administration was developed in Applesoft BASIC (Poole, McNiff & Cook, 1981). The procedure is approximately as follows: Each time the pen makes contact with the tablet or the pen is within a particular distance, the computer receives three numeric values. The first value is the horizontal (X-)coordinate, the second is the vertical (Y-)coordinate, and the third is a digit that indicates the status of the pen and of the keyboard (like "pen above or on the tablet", or "a key pressed" etc.). The sampling frequency of the tablet is set at 10 Hz and the spatial resolution at 500 lines per inch. The measurement of individual pen strokes is done in milliseconds. The data obtained per individual test sheet (all coordinates and times of pen presses are provisional data) are then recorded on disk until further analysis and processing. This sequence has to be repeated for all subtasks, i.e. all test sheets. Additional software was developed for comprehensible instructions for the investigator via the computer monitor and an explanation for the patient. The instructions are similar to those given in the respective paper and pencil test (Brand & Jolles, 1987).

Q	4	F	K	Z	G	B	D	P	9	X	N
K	R	D	3	B	F	Q	X	Z	N	R	G
5	K	X	G	9	Z	R	S	3	F	Q	B
Z	S	R	F	X	Q	4	B	G	K	R	P
F	G	P	D	S	K	B	N	P	R	Z	G
5	K	N	S	4	D	3	R	Z	B	9	F
D	R	4	Q	N	B	Z	X	F	3	S	G
N	G	P	B	S	5	Q	D	X	P	Z	F
S	5	F	D	Z	N	X	K	3	Q	R	5
Z	X	Q	S	F	D	4	N	P	G	9	K
G	P	N	9	Q	K	B	S	5	D	X	P
3	Q	X	K	R	9	D	P	N	B	4	S

Figure 1. One of the 8 different test forms of the Utrecht Memory Comparison Task (UMCT). The reproduction ratio is 1 : 2. The subject is asked to search for the targets 5-3-4-9, and to mark them with a pencil. In the other 7 test forms the positive items are the "%"-sign or 1 through 4 digits or letters.

After completion of the total test the provisional data are further analysed in a 'retrieval' program. This program allows for an exact temporal and spatial analysis of individual responses. To this end, every seperate stimulus and its exact place on the test sheet were defined.

The nature of the software that has been written for the implementation of the UMCT is such that other paper and pencil tests can be administered with only minor modifications of this program: only a separate test-specific 'retrieval' program will have to be written.

3. *Experiment 1*

3.1. *Introduction*

The first experiment was designed to investigate whether the application of the test via the XY tablet yields the same results as in the original paper and pencil test. Two parallel versions of the UMCT were administered to 20 male students in a 2 x 2 design. All subjects received both versions of the task, one by means of the tablet, one in the usual way. The instructions for the two versions were identical. This always implied the marking of those stimuli on the test sheet that belonged to the memory set (1, 2, 3 or 4 items) that had been presented immediately preceeding the test form.

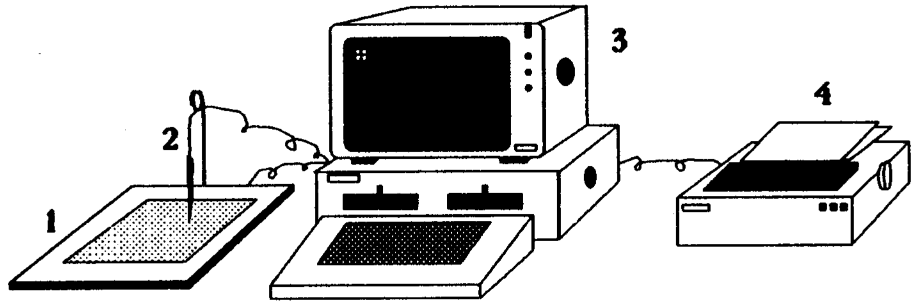

Figure 2. Schematic representation of the hardware configuration: 1. XY tablet. 2. pen. 3. microcomputer. 4. matrix printer.

3.2. Results

The results are represented graphically in Figure 3. A three-way ANOVA with test version, nature of stimuli (digits or letters), and order of test administration (paper and pencil version or computerized version first) on the intercepts of the time-setsize function did not show a significant effect of task version. This means that the computer version did not differ significantly as far as the total time needed for each test sheet is concerned. The computer version was done somewhat faster, but the difference was not significant ($F(1,18)=3.6$, $p=0.069$). A similar ANOVA was done over the slopes of the time/setsize functions. Again, task version had no effect ($F(1,18)$, ns). Moreover, these functions show the same very high linearity on both versions for those parts of the test in which letters are used as targets: there was a mean linearity of 0.95 for the paper and pencil version and 0.96 for the computerized version.

There is only a very weak correlation between the performance on both tests. This finding can most probably be explained by the fact that the experiment was carried out with students as subjects. The ceiling effect prevents the correlation as the performances show too little dispersion to correlate to any significant degree. In addition, the fact that all subjects perform two versions of the task may lead to learning effects and thus prevent the correlation.

The non-existence of visual feedback appears to have important consequences for the accuracy. The number of false negatives on the tablet version was four times higher than in the original version. In addition, averaged over all subjects, 3.5 false positives were given in the tablet version, whereas none occurred in the original version. Finally, perseverations did occur: as an average on all 9 sheets, each subject marked 35 targets more than once, whereas mistakes of this kind did not occur when a normal pencil was used. The cause of this phenomenon is not clear. It is certainly possible to make few or no mistakes on the computer version as is shown by several subjects who have high performance. The perseverations and the high number of errors may thus have to do with both the lack of visual feedback and individual variables.

Rather strong negative correlations were found between the number of perseverations and the average response time. An average correlation of -0.71 was obtained ($t=5.82$, $p<0.001$). This would justify the conclusion that speed-accuracy trade off is involved (Fitts and Radford, 1966). Apparently the instruction to work as fast as one can but at the same time to make as little mistakes as possible is differently interpreted.

The registration of individual responses makes it possible to calculate the inter-response times as a function of target distance, i.e. the number of distractors between two targets. For those test sheets for which a memory set of 4 letters was used, a 1-intercept of 732 ms and a slope of 215 ms were obtained. This possibility has far-reaching theoretical consequences. For instance, when extrapolating this function to a 0-intercept, a value of 517 ms is obtained. Within Sternberg's model, this is an estimation of the duration of the processing of one stimulus, without memory comparison. Further research into this issue will be discussed in future studies.

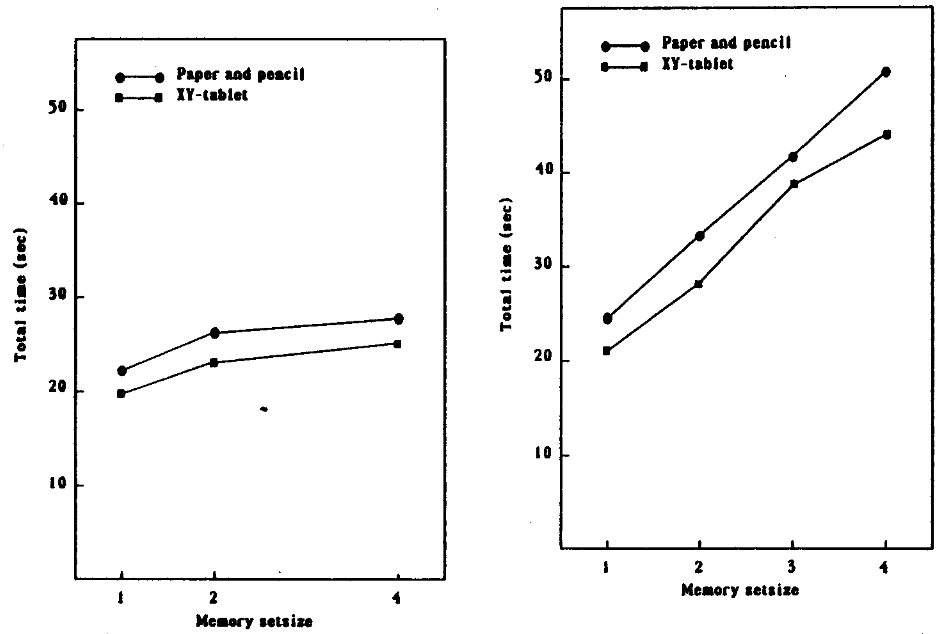

Figure 3. Mean results (total times) as a function of setsize, with digits (left) and letters (right) as memory set.

4. *Experiment 2*

4.1. *Introduction*

A model of 'serial exhaustive scanning' like the one proposed by Sternberg, predicts that a negative response takes somewhat longer than a positive response. Sternberg himself found a difference of about 45 ms per stimulus. Under the assumption that this model is valid, the influence of the factors 'the nature of the targets' and 'memory set size' should not interact with the nature of the response. Interaction between these factors would imply that the different stages of Sternberg's model would not be independent. As a consequence, such interactions would provide an argument against the model. This issue was investigated in the second experiment. The effects of the nature of the response (positive or negative) on the inter-response times was studied under various conditions. The inter-response time was operationally defined as the time elapsing between two pen strokes.

There was one experimental group which consisted of 8 subjects, all male students. Only the computer version was done. Targets had to be marked (a stroke through the item) and the distractors had to be 'underlined' (a stroke underneath the item). Each test sheet contained the usual 24 targets and 120 distractors.

4.2. *Results*

The average inter-response times of positive and negative responses are represented in Figure 4. As is evident from the figure, negative responses seem to take less time than positive responses. These differences are significant. Two two-way ANOVA's with type of response (positive or negative) and type of stimulus (digits or letters) were carried out on the slopes and intercepts of the inter-response time as a function of setsize. There was a significant effect of type of response on the slopes: $F(1,7)=39.56$, $p<0.001$. The slopes for positive responses were somewhat steeper than those for negative responses. Also, the intercepts for positive responses were about 16 msec. higher than those for negative responses ($F(1,7)=33.32$, $p<0.001$).

In the present experiment there were five times as many distractors, as usual for the UMCT, whereas in Sternberg's experiments the target/distractor ratio was 1/1. The differences may thus be due to frequency differences

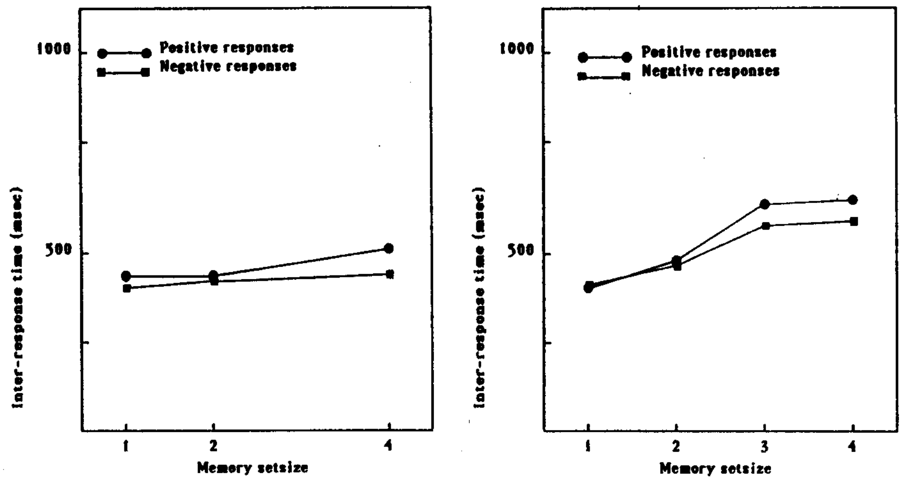

Figure 4. Mean results (positive and negative response times) as a function of setsize, with digits (left) and letters (right) as memory set.

between targets and distractors. When a subject has to switch from a negative response to a positive response, he/she has to interrupt a certain pattern of movements. It is likely that this takes some time. Therefore, the possibility cannot be ruled out that a different outcome would have been obtained with an equal amount of targets and distractors. This will be explored in a follow-up study.

Another finding which is in contradiction to Sternberg's model relates to the size and nature of the memory set. The effects of these factors interacted significantly with those of the nature of the response. As was stated in paragraph 4.1., a serial model of processing like Sternberg's prohibits interactions of any kind, as an interaction would mean that at least two factors are not independent. Consequently, these findings run counter to Sternberg's model.

5. *Discussion*

The two experiments reported in the present study indicate that the performance on the UMCT is similar in the XY tablet version and in the normal version. The time needed for each stimulus sheet and the time as a function of setsize are comparable on both versions.

Response times with the computer version are systematically (however not significantly) shorter. It is possible that the subject is inclined to work less accurately with a non-writing pen and to touch the paper more superficially. This interpretation is strengthened by observations done during the administration of the test. Furthermore, together with the absence of visual feedback, reduced accuracy would explain the much larger number of errors made on the computer version.

It was remarkable that in spite of the instructions to mark the positive items by a stroke, many subjects tended to respond by touching them only. It goes without saying that making a stroke takes more time than merely touching, and this probably explains the difference in time. This issue is subject to further experimentation, with a pen that leaves a normal, visible trace. This is quite feasible, as is clear from writing research carried out with such a device (Maarse, Schomaker & Teulings, 1986).

With respect to the experience with the use of the XY tablet, and particularly the Apple Graphics Tablet, the following findings may be relevant: No difficulties were experienced with test administration. There were, however, some technical problems. The sampling frequency of the tablet which is, in combination with a BASIC program, relatively low, renders the accurate interpretation of the provisional data by the retrieval program rather difficult. It is now considered to develop an assembler routine that can speed up the sampling frequency considerably. Such a routine should be universally applicable in view of other paper and pencil tests, which are to be implemented on the tablet.

It became clear that the pen, which was purchased with the tablet, was not sufficiently solid to meet the requirements. The pen had to be replaced as it appeared that the pen pressure could not -or not with sufficient reliability- be ascertained. An additional disadvantage was that the subjects had to be instructed not to press the pen too weakly, since the built-in contact requires a certain minimum pressure.

The use of a pen that leaves an ink trace as used in the study of Maarse, Schomaker, and Teulings (1986), would allow a number of tests to resemble the real paper and pencil way of testing even more. Moreover it would facilitate the investigation of the effects of visual feedback as it would bring visual feedback under experimental control. The use of an inking pen has more 'ecological validity'. Writing is a more common activity than touching symbols or drawing invisible lines.

The method under discussion seems to yield a very suitable synthesis between the short duration and the simplicity of the paper and pencil test on the one hand, and the accuracy and reliability of the computer controlled assessment on the other. The use of the tablet makes the test hardly any longer than the original paper and pencil test. Moreover, the testing can be done on the same sheets, whereas all instructions can be identical, certainly with a pen that writes. In this way the test can remain comprehensible and conform more to everyday activities than for instance a monitor test which requires the handling of response keys. From the experiments discussed, it may be clear that considerably more variables can be measured than before.

Thus, application of the XY tablet with standard paper and pencil tests seems possible with the procedure described in the present paper. Implementation of other tests, such as the Trail Making Test, the Gibson Spiral Maze, the Road Map Test, Embedded Figures Tests, etc. (see Lezak, 1983) will be the subject of following studies.

Computer-aided assessment may thus add new possibilities to neuropsychological assessment and enable the use of information paradigms in the clinic (Jolles, 1985).

References

Brand, N., & Jolles, J. (1986). Information processing in depression and anxiety. *Psychological Medicine, 17,* 145-153.

Fitts, P.M., & Radford, B.K. (1966). Information capacity of discrete motor responses under different cognitive sets. *Journal of Experimental Psychology, 71,* 475-482.

Jolles, J. (1985). Early diagnosis of dementia: possible contributions from neuropsychology. In W.H. Gispen & J. Traben (Eds.), *Aging of the brain.* Berlin: Springer-Verlag.

Jolles, J., Hijman, R., & Gaillard, A.W.K. (1982). Geheugenstoornissen en informatieverwerking [Memory disorders and information processing]. *De Psycholoog, 17,* 487.

Lezak, M.D. (1983). *Neuropsychological Assessment.* New York-Oxford: Oxford University Press.

Maarse, F.J., Schomaker, L.R.B., & Teulings, H.L. (1986). Kenmerkende verschillen in individueel schrijfgedrag. Automatische identificatie van schrijvers [Discriminative features in individual writing behavior]. *Nederlands tijdschrift voor de Psychologie, 41,* 41-47.

Poole, L., McNiff, M., & Cook, S. (1981). *Apple II User's Guide.* Berkeley: Osborne/McGraw-Hill.

Sternberg, S. (1969). The Discovery of Processing Stages: Extensions of Donders' Method. *Acta Psychologica, 30,* 276-315.

Sternberg, S. (1975). Memory Scanning: New Findings and Current Controversies. Quarterly *Journal of Experimentul Psychology, 27,* 1-32.

Thurstone, L.L. (1944). *A factorial study of perception.* Chicago: University of Chicago Press.